

협동 센서 융합 기반 화자 성별 분류를 위한 무선 센서네트워크 개발

A Development of Wireless Sensor Networks for Collaborative Sensor Fusion Based Speaker Gender Classification

권호민*

Ho-min Kwon*

요약

본 논문에서는 무선센서네트워크에서 이루어지는 협동적 센서융합을 이용한 화자성별분류를 제안하였다. 센서노드들은 BER(Band Energy Ratio) 기반 음성활동검출을 수행함으로써 불필요한 입력 데이터는 제거하고 관련성이 높은 데이터만을 처리 및 경판정한다. 개별적 센서노드에서 생성된 경판정 값들은 융합센터로 송신되고 전역적 결정 융합을 구축하기 때문에 전력 소모를 줄이고 네트워크 자원을 절약한다. 화자성별분류를 위한 센서융합기법으로써 베이시안(Bayesian) 센서융합 및 전역적 가중결정융합기법들이 제안되었다. 베이시안 센서융합의 경우, 배치되는 센서노드 수 변화에 따른 ROC(Receiver Operating Characteristic) 커브의 움직임을 통해 개별 센서노드 레벨에서 얻어진 경판정 값들을 처리하고 최적의 분류 융합을 결정한다. 전역적 결정을 위한 가중치로써 BER 및 MCL(Mutual Confidence Level)을 채택하여 개별적 지역 경판정 값들을 효율적으로 결합 및 융합시킨다. 센서 노드의 수가 증가함에 따라 분류화 성능이 개선되어졌으며 특히 낮은 SNR(Signal to Noise Ratio) 환경에서 성능 개선폭이 더 높게 나타남을 실험적으로 확인하였다.

Abstract

In this paper, we develop a speaker gender classification technique using collaborative sensor fusion for use in a wireless sensor network. The distributed sensor nodes remove the unwanted input data using the BER(Band Energy Ration) based voice activity detection, process only the relevant data, and transmit the hard labeled decisions to the fusion center where a global decision fusion is carried out. This takes advantages of power consumption and network resource management. The Bayesian sensor fusion and the global weighting decision fusion methods are proposed to achieve the gender classification. As the number of the sensor nodes varies, the Bayesian sensor fusion yields the best classification accuracy using the optimal operating points of the ROC(Receiver Operating Characteristic) curves. For the weights used in the global decision fusion, the BER and MCL(Mutual Confidence Level) are employed to effectively combined at the fusion center. The simulation results show that as the number of the sensor nodes increases, the classification accuracy was even more improved in the low SNR(Signal to Noise Ration) condition.

Keywords : Wireless Sensor Network, Bayesian Fusion, Decision Fusion, Gender Classification

I. 서 론

신호처리 기술을 기반으로 한 집적된 무선 송수신기와 하드웨어 시스템의 출현은 무선센서네트워크 분야에서 여러 진보된 시스템 어플리케이션 개발을 가능하게 하였다[1]. 최근 협동신호처리 기법들은 무선센서네트워크의 분산된 센싱 및 분류화 프레임워크(framework)에서 여러 이점들을 보여주고 있다. 센서네트워크 영역에 위치된 관심 물체들은 분산된 센서들을 이용, 센서퓨전과 같은 협동알고리즘처리를 기반으로 더 중요한 정보들을 해석해낼 수 있다. 이런 패러다임은 종전의 단일 센서가 모든 입력 데이터를 수집하고 해석하는 것과는 차이가 있다. 이러한 협동신호처리기법은 환경적 소리 관찰과 측정, 이동중인 차량분류,

물체추적 등에 적용되어지고 있다 [2][3]. 이런 분야에서 음향센서는 특정 환경에서 다양한 정보를 제공하기 때문에 그 중요성이 더해진다. 특히 사람의 음성의 경우 화자의 성별, 나이, 감정상태, 대화 내용 등 중요한 화자정보를 온전히 전달해주기 때문에 가장 중요한 음향적 단서가 된다. 음향현상분석의 이론적인 측면은 음원의 분리, 개별적 음향기호 분석, 창의적 매개변수화, 개인화 정보검출 및 협동센서네트워크 구축을 요구한다 [4][5][6]. 시스템구현 측면에서도 하드웨어 자원의 공유, 고정소수점 처리, 자동 절전모드, 실시간 동작을 위한 노드들간의 공유 처리 등 여러 난해한 문제점들을 포함한다.

무선센서네트워크에 적용되는 융합기법은 다수결 원칙(Majority Voting)을 이용한 단순한 방법에서부터 특정 어플리케이션에 최적화된 방법에 걸쳐 매우 다양하게 발전되어 왔다 [7]. 가장 일반적인 방법은 네트워크의 기지국에서 모든 센서노드로부터 모든 데이터를 수집하는 것이다. 이런 경우 센서 노드들은 관측해야 할 신호에 대한 사전 정보를 가지고 있지 않다. 따라서 최적 해결 수

* 삼성전자

투고 일자 : 2011. 1. 31 수정완료일자 : 2011. 4. 4

제작확정일자 : 2011. 4. 30

식은 MSC(magnitude-squared coherence) 추정을 통해서 결정될 수 있다. 두 개의 센서를 위한 테스트 통계치는 식 (1)과 같다.

$$\Gamma^2(x_1, x_2) = \frac{|\langle x_1, x_2 \rangle|^2}{\|x_1\|^2 \|x_2\|^2} \quad (1)$$

여기서 x_1 과 x_2 는 센서 1과 센서 2에서 수집된 신호를 나타낸다. 이 테스트 통계치는 센서노드들이 수집된 모든 데이터를 MSC 추정 및 판단을 위해 기지국으로 전송되어야 한다. 이때 판측신호의 존재 여부에 무관하게 수행된다. 이것은 데이터 전송비용과 전력소모 측면에서 살펴보면 효율성이 떨어진다. 고효율 측면에서 접근 가능한 방법은 미리 규정된 결정규칙을 바탕으로 각각의 센서가 개별적 판단결과를 산출하고 그 판단된 데이터를 전송하는 것이다. 이런 경우 테스트 통계치 Λ 는 식 (2)와 같다.

$$\Lambda = \prod_{d_m=1} P_d^m \prod_{f_a} \prod_{d_m=0} \frac{1-P_d^m}{1-P_f^m} \quad (2)$$

여기서 P_d^m 와 P_f^m 는 센서 m 에서의 검출 확률(probability of detection)과 오검출 확률(probability of false alarm)이다. 최종 결정은 고정된 상수인 P_d^m 와 P_f^m 값을 이용하여, 개별판단 값을 가중치를 적용하여 얻어진다. 여기서의 융합 규칙은 각 센서노드들과 기지국이 확률값들을 사전에 알고 있다고 가정한다. 만약 다변량 입력에 따라 P_d^m 와 P_f^m 가 변한다면 이러한 융합방법도 실제 시스템에 적용하기에는 무리가 있다. 따라서 협동센싱 테크닉은 노드레벨에서 얻어진 개별적 정보들이 분석되거나 시스템 성능 개선을 위해 서로 다른 센서노드들간의 정보들이 지능적으로 결합해질 때 가장 효율적인 결과를 낼 수 있다.

본 논문에서는 분산된 개별 센서들에서 획득되는 정보들을 이용하는 협동적 센싱 및 융합 프레임워크를 제안한다. 네트워크에 분산되어진 센서노드들은 음성활동검출을 이용하여 불필요한 입력데이터를 제거하고 개별적 판단을 수행, 전력소모를 줄이고 네트워크 자원을 절약할 수 있다. 일반적으로 무선센서네트워크에서 이루어진 음성센싱기법들은 단일화자에 적용된 화자위치검출 및 추적에 이용된다. 본 논문은 화자성별분류를 위해 센서융합에 있어 베이시안(Bayesian) 센서융합 및 전역적 가중융합결정기법들을 제안하였다. 균일하게 개별 정보들을 처리하는 종래의 융합기법들과 달리 본 논문에서는 가변적 입력 노이즈에 따른 관심있는 신호의 충실도를 주파수영역에서 개별적으로 분석하여 효과적으로 가중시킴으로써 전역적 결정(global decision)을 얻는다. 본 논문에서 언급되는 ROC(Receiver Operating Characteristic) 커브들은 검출 확률 및 오검출 확률 관점에서 성별검출성능을 나타낸다. 특정공간에서 센서 수가 증가함에 따라 화자성별검출기의 성능은 개선되어 지고 제안된 프레임워크는 불필요한 입력 데이터를 제거, 협동적으로 센서융합을 수행함으로써 종래의 융합방법들보다 다양하게 추가되는 백색 잡음 환경에서강인한 성능을 보여준다.

II. 협동적인 센싱 및 융합

2.1 음향현상 분석을 위한 협동적 센싱 프레임워크

무선센서네트워크에서 전력소비를 줄이고, 효율적인 운영을 위

해서는 센서노드에서 적절한 정보처리가 이루어지고 융합센터 혹은 기지국에서 협동적 융합을 이루는 것이 도움이 된다. 기존의 융합 결정법은 특정 가중치를 이용하여 개별적 판단들이 결합되는 형태를 사용한다. 예를 들어, Neyman-Pearson 방법은 공간적, 시간적 독립성을 가정하고 개별 판단값들의 가중된 합을 고정된 임계치와 비교하여 전역적 결정을 한다 [8]. 반면 통계적 모델들은 각 센서노드에서 MAP(Maximum a Posteriori) 추정하고 센서 융합을 함으로써 무선통신에 의한 더 많은 전력소비를 요구한다. 소비되는 전송전력은 개별적 판단값들을 전송하는 센서 노드수에 비례하기 때문이다. 본 논문에서는 각각의 센서노드에서 부분적으로 처리된 정보들을 경판정(Hard Labeling)하여 융합 센터로 보냄으로써 전송전력을 절약한다.

본 논문에서는 제안하는 무선협동센서네트워크는 개별의 센서노드들이 특정 센싱함수를 수행하고 다른 센서노드들 및 융합센터와 수집된 정보를 공유한다. 임의의 시간대에 센서노드들은 전역적 판단이 가능한 모든 정보들을 가지고 있게 된다. 시스템 자원 효율을 고려할 경우, 소모 전력이 제한적인 어플리케이션의 경우에는 센싱처리단계에서 관련성이 낮은 데이터들을 제거한다. 각 센서 노드에서 동작하는 음성 활동 검출은 수행해야 할 task와 관련된 정보들만을 처리한다. 이러한 접근방법은 분류화 처리에 있어 불필요한 입력 데이터(비음성 데이터)를 제거하는데 효과적이다. 또한 성별 검출에 있어 음성 활동 구간의 피치(Pitch)와 캡스터럼(Cepstrum) 계수들이 특징 벡터로 사용된다.

2.2 음성 활동 검출

음성신호는 특정 주파수 영역에서 비교적 많은 에너지를 포함하기 때문에 power spectrum의 Band Energy Ratio(BER)을 이용하여 검출된다. BER은 식 (3)과 같다.

$$v_{BER}(i) = \frac{\sum_{f_p}^{f_q} S_i(f)}{\sum_{f=1}^M S_i(f)}, \quad 0 < v_{BER}(i) < 1, \quad (3)$$

여기서 f_p 와 f_q 는 선택된 주파수 영역의 시작과 종료 인덱스를 말하며, $S_i(f)$ 는 i 번째 프레임의 FFT 결과를 말한다. 여기서는 16KHz 샘플링 주파수, M=256일 때 125Hz와 2KHz에 상응하는 $f_p=4$ 와 $f_q=64$ 를 사용했다. 계산된 BER 결과는 변화하는 SNR(Signal to Noise Ratio) 환경에서 적응적 신호검출을 위해 이용된다. 초기화 작업으로 입력음성신호 10 프레임의 획득된 BER의 평균 μ 및 분산 σ 을 구한다. 목표 신호는 BER이 임계치를 넘었을 때 검출되며 임계치는 아래의 식 (4)와 같이 주어진다.

$$\delta_{target} = \mu_i + \tau\theta \quad (4)$$

여기서 τ 는 임계치 매개변수이며 θ 는 고정된 상수값이다. 이 수식은 식 (5)와 같이 주변 노이즈 레벨의 변화에 적응하기 위해 목표 신호가 검출 되지 않을 시 조건적으로 업데이트된다.

$$\mu_i = \alpha v_{BER}(i) + (1-\alpha) \mu_{i-1}, \quad 0 < \alpha < 1 \quad (5)$$

여기서 α 는 적응 매개변수이다. 예상치 못한 노이즈 조건들로

인해 급격한 BER의 변화를 피하기 위해서 α 값은 낮아야 한다. 그림 1은 $\alpha=0.1$ 이고 $\tau=2.5$ 일 때 백색 가우시안 노이즈를 이용 두 개의 다른 SNR 환경에서 ITU G.729 [10] 음성 활동 검출 기법과 BER 기반 음성 활동 검출 기법의 성능을 보여준다. 그림 1(a)는

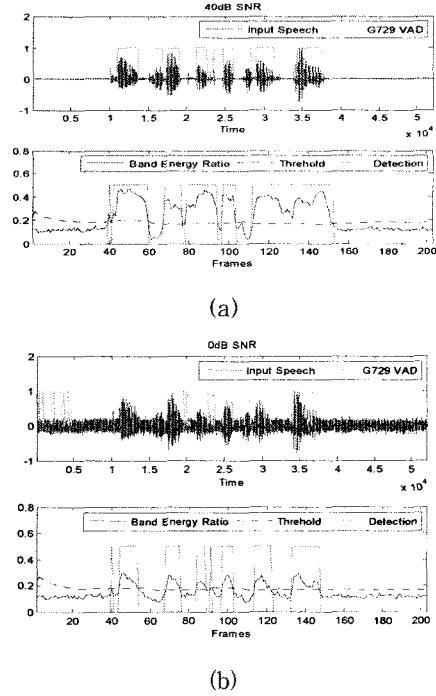


그림 1. ITU G.729 검출기법과 BER기반 적용 검출기법 비교
(a) 40dB SNR 환경 (b) 0dB SNR 환경.

Fig. 1. The comparison of ITU G.729 VAD and the BER based VAD (a) 40dB SNR (b) 0dB SNR.

두 개의 방법이 모두 40dB SNR 환경에서 좋은 검출 성능을 보여준다. 그림 1(b)에서는 BER 기반 VAD 기법이 0dB SNR에서 G.729 VAD 기법보다 낮은 오검출율을 보이며 노이즈보다 높은 신호 레벨 영역에서 목표신호가 검출되는 것을 나타낸다.

2.3 개별 센서 노드 레벨에서의 판단 기법

추가적인 전력소모를 줄이기 위한 기법으로 분산된 센서노드들이 모든 데이터를 송신하는 대신 개별적으로 처리된 판단 결과값을 기지국으로 송신한다. 노드 레벨에서 2.2절에 기술된 음성활동 검출기는 유효한 음성신호를 검출하고 이 구간동안 음성특징들을 검출한다. 검출된 특징들은 개별적 판단 결과값을 만들기 위해 사전 학습된 분류기의 입력으로 사용된다. 그림 2은 개별 센서노드에서 이루어지는 개별적 화자성별판단기법을 나타낸다. 각각의 센서노드에서 수집되고 처리된 음성특징들을 바탕으로 얻어진 개별적 판단값이 융합센터로 전송된다. 본 논문은 이를 위해 베이시안 센서 융합과 전역적 가중치결정기법을 제안한다. 베이시안 융합에 있어 성별의 사후확률(Probability of Posterior)이 개별적으로 계산되고 융합센터에서는 우도 비율(likelihood ratio)로 최적의 결정을 도출해낸다. 전역적 가중치결정에 있어 각각의 경판정된 결정들이 SNR 추정과 상호신뢰도레벨 (MCL: Mutual Confidence L-

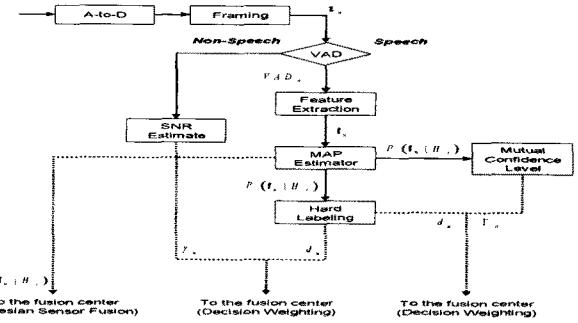


그림 2. 무선 센서 네트워크에서의 성별 분류를 위한 개별적 판단 기법.

Fig. 2. The local decision process of gender classification for used in a wireless sensor network.

evel)로 가중처리된다. 센서 u 는 $u \in \{1, 2, 3, \dots, U\}$ 의 집합 구성을 가지며 u 센서노드에서의 음성 활동 검출 VAD_u 와 입력 신호의 프레임 x_u 를 정의한다. 2.1 절에서 기술된 음성 활동 구간을 검출하고 일련의 특징벡터 f_u 를 추출해내는 특징 추출 함수 $F()$ 는 다음과 같이 정의된다.

$$f_u = F(x_u, VAD_u) \quad (6)$$

본 논문에서는 화자성별 분류를 위해 신뢰할만한 성능을 보여주는 MFCCs(Mel-Frequency Cepstrum Coefficients) [11] 와 퍼치, 그리고 이 값들의 델타 값을 이용하여 가우시안 혼합 모델(GMM : Gaussian Mixture Model)로 분류한다 [12]. 본 논문에서는 센서 노드에서 추출되는 음성 특징벡터는 pitch + Δ pitch + 14차 MFCCs + Δ MFCCs 형태로 이루어져 있으며 MAP를 이용, GMM을 학습하였다. 두 가지의 가설 H_0 (여성) 과 H_1 (남성) 하에 사후 확률은 $P(f_u|H_0)$ 와 $P(f_u|H_1)$ 로 표현한다. 이 확률값들은 융합센터에서 베이시안 센서융합된다. 그리고 개별적 판단은 아래의 식 (7)과 같이 계산되고 경판정된다.

$$d_u = \begin{cases} 1, & \text{if } P(f_u|H_0) \geq P(f_u|H_1) \\ -1, & \text{otherwise} \end{cases} \quad (7)$$

전역적 결정가중융합을 위해 본 논문에서는 수집된 데이터와 깊은 연관성을 가지는 값으로 SNR γ_u 와 MCL σ_u 을 제안한다. 입력 신호의 SNR을 가중치로 적용할 경우 개별적 판단의 신뢰성 판단에 도움이 되지만 자연환경에서 발생하는 노이즈는 예측 불가능하기 때문에 이 값을 사용하기는 불가능하다. 이를 대체하기 위해서 음성신호의 특정 주파수대의 에너지를 비교, 노이즈에 대한 상대적 신호세기를 나타내는 식 3에 기술된 BER을 사용한다. 한편 사후확률 $P(f_u|H_i), i \in \{0, 1\}$ 은 경판정된 개별적 판단들의 신뢰도를 반영한다. 사후 확률값이 높을수록 그에 대응하는 개별적 판단의 신뢰도는 증가한다. 따라서 각각의 개별적 판단값들을 성별 가설들로부터 얻어지는 MCL로 가중시킬 수 있다. MCL은 다음의 식 (8)과 같이 정의된다.

$$\Gamma_u = |P(f_u|H_0) - P(f_u|H_1)| \quad (8)$$

2.4 베이시안 센서 융합

u 번째 센서 노드에서 MAP 추정과 가설 $H_i, i \in \{0, 1\}$ 이 주

이겼을 때 베이시안 위험 [13]은 식 (9)와 같이 주어진다.

$$R = \sum_{i=0}^1 \sum_{j=0}^1 C_{ij} P(H_j) P(f_u | H_j) \quad (9)$$

본 논문에서는 두 개의 가설만 고려하므로 식 (9)에 따른 C_{00} , C_{10} , C_{11} , C_{01} 은 4개의 분류확률에 따르는 비용이다. 따라서, 총 판측 Z 은 두 개의 영역 Z_0 와 Z_1 으로 구분된다. 판측 결과 $\int_Z P(f_u | H_0) df_u = \int_Z P(f_u | H_1) df_u$ =로부터 아래의 식 (10)의 베이시안 위험을 유도해 낼 수 있다.

$$\begin{aligned} R &= C_{10} P(H_0) + C_{11} P(H_1) \\ &+ \int_Z P(H_1) (C_{01} - C_{11}) P(f_u | H_1) df_u \\ &- \int_Z P(H_0) (C_{10} - C_{00}) P(f_u | H_0) df_u \end{aligned} \quad (10)$$

참 분류의 비용이 거짓 분류의 비용보다 크다고 가정할 경우 (i.e., $C_{10} > C_{00}$ 그리고 $C_{11} > C_{01}$), 판단 바운더리 (boundary)는 다음의 식 (11)과 같이 정의된다.

$$P(H_1) (C_{01} - C_{11}) P(f_u | H_1) \geq P(H_0) (C_{10} - C_{00}) P(f_u | H_0) \quad (11)$$

u 번째 센서 노드에서 얻어지는 우도는 식 (12)와 같이 정의된다.

$$A_u = \frac{P(f_u | H_1)}{P(f_u | H_0)} \quad (12)$$

베이시안 센서퓨전 d_{BF}^f 은 최적의 우도판단규칙을 다음 식 (13)과 같이 만든다.

$$d_{BF}^f = \begin{cases} H_1, & \frac{P(f_1, f_2, \dots, f_u | H_1)}{P(f_1, f_2, \dots, f_u | H_0)} > TH \\ H_0, & \text{otherwise} \end{cases} \quad (13)$$

여기서 TH 는 정해진 임계값이다. f_u 의 통계적 독립을 가정할 때 식 13에 있는 결합확률은 다음 식 (14)와 같다.

$$P(f_1, f_2, \dots, f_u | H_i) = \prod_{u=1}^U P(f_u | H_i), \quad i = 0, 1 \quad (14)$$

2.5 전역적 판단 가중 기법

2.3 절에서 기술되었듯이 전역적 결정가중융합기법은 BER과 MCL을 이용한다. BER를 적용함에 있어 본 논문은 두 가지 가중 기법을 제안한다. 첫 번째 기법은 지역 판단에 상응하는 BER를 가중시키는 것 (d_{BER}^f)이며 다음의 식 (15)와 같이 정의된다.

$$d_{BER}^f = \frac{\sum_{u=1}^U v_{BER}^u d_u}{\sum_{u=1}^U v_{BER}^u}, \quad 0 < v_{BER}^u < 1 \quad (15)$$

여기서 v_{BER}^u 은 u 번째 센서 노드의 BER 가중치이며 U 는 모든 개별적 판단개수를 가리킨다. d_u 는 u 번째 센서 노드의 판단을 가리키며 다음 식 (16)과 같다.

$$d_u = \begin{cases} 1, & \text{if 여성} \\ -1, & \text{if 남성} \end{cases} \quad (16)$$

두 번째 기법은 최대 BER을 가지는 개개 다음 식 (17)과 같다.

$$d_{BER_{max}}^f = d_u^* \quad (17)$$

여기서 u^* 은 식 (18)과 같이 정의된다.

$$u^* = \arg \max v_{BER}^u \quad (18)$$

MCL을 이용한 판단가중기법 d_{MCL}^f 은 다음 식 (19)와 같이 정의된다.

$$d_{MCL}^f = \frac{\sum_{u=1}^U \Gamma_u d_u}{\sum_{u=1}^U \Gamma_u}, \quad 0 < \sigma_u < 1 \quad (19)$$

추가적으로 본 논문에서는 위에서 기술한 가중 기법들과 성능을 비교하기 위해서 기존의 융합 기법인 다수결의 원칙 결정법 d_{MV}^f 을 고려했으며 이는 다음 식 (20)과 같다.

$$d_{MV}^f = \begin{cases} 1, & \text{if } \sum_{u=1}^U d_u \geq 0 \\ -1, & \text{otherwise} \end{cases} \quad (20)$$

III. 화자 성별 분류 결과 및 성능 비교

그림 3은 음향 센서노드들의 다양한 지형적 위치를 나타낸다. 센서노드의 수는 1개에서 24개까지 총 6가지 배치를 고려하였으며 단수의 음성신호와 단수의 노이즈 신호가 임의적으로 위치한다. 반향음은 존재하지 않는다고 가정했으며 신호세기는 신호 위치와 센서 사이의 거리에 따라 -2 지수로 감쇠하도록 설정했다. TIMIT 데이터베이스 [14]를 이용하여 64 혼합수의 GMM 분류기를 학습하였다. 음성 문구마다 화자 분류를 위해 200회의 Monte-Carlo 실험이 이루어졌다. 베이시안 센서융합의 실험 통계

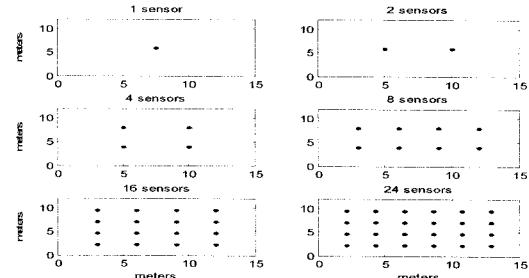


그림 3. 화자 성별 분류를 위한 음향 센서노드의 지형적 위치.

Fig. 3. The geometry of the acoustic sensors for gender classification.

표 1. 화자 분류를 위한 ROC 커브들로부터 얻어진 동작점.

Table 1. The operating points of the ROC curves for gender classification.

SNR	센서 수	여성(P_d , P_{fa})	남성(P_d , P_{fa})
20dB	1	(0.94, 0.02)	(0.98, 0.06)
	2	(0.97, 0.02)	(0.98, 0.03)
	4	(0.96, 0.01)	(0.99, 0.04)
	8	(0.96, 0.01)	(0.99, 0.04)
	16	(0.96, 0.02)	(0.98, 0.04)
	24	(0.88, 0.02)	(0.98, 0.12)
0dB	1	(0.89, 0.16)	(0.84, 0.11)

2	(0.93, 0.12)	(0.88, 0.07)
4	(0.95, 0.09)	(0.91, 0.05)
8	(0.94, 0.08)	(0.92, 0.06)
16	(0.93, 0.08)	(0.92, 0.07)
24	(0.95, 0.08)	(0.92, 0.05)

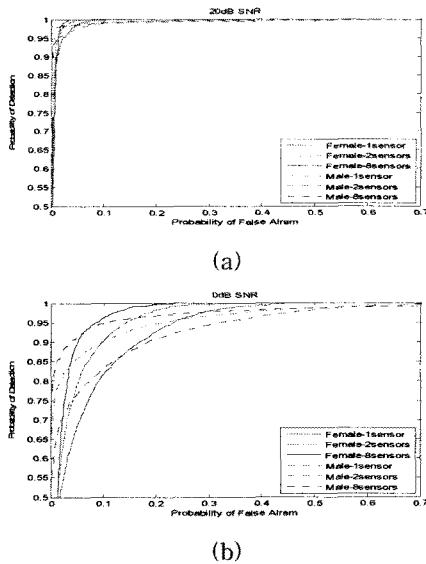


그림 4. SNR 환경에서의 ROC 커브 비교 (a) 20dB SNR환경
(b) 0dB SNR 환경.

Fig. 4. The comparison of the ROC curves (a)at 20dB SNR (b)at 0dB SNR.

치는 여성과 남성 분류를 위한 임계값으로 사용되었으며 임계값에 따라 검출 확률(P_d)과 오검출 확률(P_f)은 변하게 된다. 그에 따른 ROC 커브들을 그림 4에 나타냈다. 낮은 오검출 확률에서는 남성 음성의 검출 성능이 여성 음성보다 우수하였으며 높은 오검출 확률에선 그 반대로 동작하였다. 센서노드의 수가 증가함에 따라 성별 분류기의 성능은 0dB SNR에서보다 20dB SNR에서 더 높은 성능을 보였다. 표 1은 센서노드 수에 따른 최고의 성별분류성능

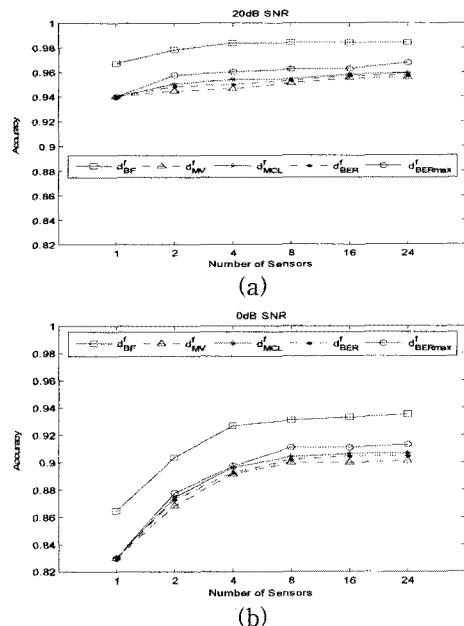


그림 5. SNR 환경에서의 센서 융합 성능 비교 (a) 20dB SNR 환경 (a) 0dB SNR 환경

Fig. 5. The performance comparison of the sensor fusion (a)at 20dB SNR (b)at 0dB SNR.

을 보여주는 동작점 (P_d , P_f)를 나타낸다. 전역적 결정가중융합 기법은 1과 -1사이에서 판단값이 형성되므로 단일 임계치 ($TH=0$)으로 동작한다. 그림 5는 20dB 및 0dB SNR 환경에 있어 베이시안 센서융합 (d_{BF}^f)과 4가지 전역적 결정가중융합기법들 ($d_{MV}^f, d_{MCL}^f, d_{BER}^f, d_{BER\max}^f$)의 분류정확도를 보여준다. 모든 경우에 있어 d_{BF}^f 가 가장 좋은 성능을 나타내었다. 센서노드 수가 증가함에 따라 분류 정확도는 20dB SNR에서 1.6%, 0dB SNR에서 6.8% 개선되었다. 이는 d_{BF}^f 가 경판정된 판단값들 보다 사후 확률로 더 많은 정보를 제공하기 때문이다. $d_{BER\max}^f$ 가 두 번째로 좋은 성능을 보였다. 센서노드 수가 변화함에 따라, 분류 정확도는 20dB SNR에서 2.1%, 0dB SNR에서 8.4% 개선되었다. 이는 $d_{BER\max}^f$ 가 가장 높은 BER이 포함된 센서 개별 판단값이 다른 가중 기법들보다 더 신뢰할 수 있고 연관성 높은 가중치를 내포하기 때문이다. d_{MCL}^f 은 d_{BER}^f 과 d_{MV}^f 에 비해 조금 나은 성능을 보였다. d_{MV}^f 가 모든 개별적 판단에 동일한 중요도를 부여하는 만큼 대비해서 d_{MCL}^f 과 d_{BER}^f 은 강인한 가중치 효과를 보여주지는 않았다. 음성신호와 노이즈신호는 임의적으로 무선센서네트워크 내에 위치하므로 센서노드들의 위치는 센서융합성능에 큰 영향을 미치지 않는다. 센서노드의 수가 증가함에 따라 분류성능의 증가폭은 감소했다. 이는 더 많은 센서노드가 네트워크에 포함될수록 얻어지는 새로운 정보의 양은 감소하기 때문이다. 종합적으로 고려할 때 분산 되어진 센서노드들의 개별적 판단값의 융합은 백색잡음 강도가 증가할수록 분류 성능을 개선하는데 장점이 있다. 비록 베이시안 센서융합기법이 가장 좋은 결과를 산출했지만 BER을 이용한 개별적 판단 가중 기법도 성능을 개선하는데 이점이 있음을 확인할 수 있었다.

IV. 결론

무선 센서 네트워크에서 협동적 센싱과 융합을 위해 분산된 센서노드들에서 측정된 개별적 정보들이 효과적으로 가중되어 결합되었다. 본 논문에서는 화자성별분류를 위한 베이시안 센서융합기법이 적용 되었으며 전역적결정가중기법이 개발되었다. 베이시안 센서융합기법에 있어서, ROC 커브를 통해서 최적의 동작점을 확인하였고, 화자성별분류에 있어서 최고의 성능을 나타내는 것을 확인하였다. 전역적 결정 가중치로써 BER과 MCL을 채용, 결정융합기법을 도출해냈다. 융합기법들은 기쪽존의 융합기법인 다수결의 원칙과 그 성능을 비교하였다. 본 논문에서 기술된 융합 기법들은 낮은 SNR 환경에서 분류 성능이 개선됨을 확인할 수 있었다. 본 논문에서 제안된 무선센서네트워크 기반 센싱 프레임워크는 양/질적인 관점에서 중요한 화자정보를 추출해내고 효과적으로 처리하였다. 이는 향후 화자 정보를 무선센서네

트워크에서 분석하기 위해 전력소모감소와 효율적인 신호처리 연산 등의 관점에서 중요한 지표로 활용될 수 있음을 기대할 수 있다.

참고 문헌

- [1] 노광현, 이병복, 박애순, “유비쿼터스 홈네트워크를 위한 LoanRF 디바이스 기반의 센서 네트워크 설계 및 응용,” 한국신호처리시스템논문지, 제7권, 제3호, pp. 87-94, 7. 2006.
- [2] G. Wichern, H. Thornburg, and A. Spanias, “Multi-channel audio segmentation for continuous observation and archival of large spaces,” *Proc. of ICASSP'09*, pp. 237-240, 2009.
- [3] A. Arora, P. Dutta, S. Bapat, V. Kulathumani, H. Zhang, V. Naik, V. Mittal, H. Cao, M. Demirbas, M. Gouda, Y. Choi, T. Herman, S. Kulkarni, U. Arumugam, M. Nesterenko, A. Vora, and M. Miyashita, “A line in the sand: A wireless sensor network for target detection, classification, and tracking,” *Computer Networks*, Vol. 46, no. 5, pp. 605-634, Dec. 2004.
- [4] A.S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT press, 1994.
- [5] V. Berisha, H. Kwon, and A. Spanias, “Real-time acoustic sensing using wireless sensor motes”, *Proc. of ISCAS'06*, pp. 847-850, 2006.
- [6] L. Lu, H.-J. Zhang, H. Jiang, “Content analysis for audio classification and segmentation,” *IEEE Trans. on Speech and Audio Processing*, Vol. 10, pp. 504-516, 2002.
- [7] D. Hall and S. McMullen, *Mathematical Techniques in Multisensor Data Fusion*, Artech House, 2004.
- [8] P. K. Varshney, *Distributed Detection and Data Fusion*, Springer-Verlag, New York, 1997.
- [9] A. Benyassine, E. Shlomot, and H. Su, “ITU-T recommendation G.729, annex B, a silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications,” *IEEE Communication Magazine*, pp. 64 - 72, 1997.
- [10] S.B. Davis and P. Mermelstein, “Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences”, *IEEE Trans. on Acoust., Speech and Signal Processing*, Vol. 28, pp. 357 - 366, 1980.
- [11] J. A. Bilmes, “A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models,” Technical Report, University of Berkeley, ICSI-TR-97-021, 1997.
- [12] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, Wiley Interscience, 2001.
- [13] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, “The DARPA TIMIT acoustic-phonetic continuous speech corpus CD ROM,” Tech. Rep. NISTIR 4930 / NTIS Order No. PB93-173938, National Institute of Standards and Technology, Gaithersburgh, Md, USA, Feb. 1993.



권 호민(Ho-Min Kwon)

2002년 2월 동아대 전자공학과(공학사)

2004년 2월 동아대 전자공학과(공학석사)

2009년 12월 애리조나 주립대 전기공학과(공학박사)

2010년 1월 ~ 2010년 10월 (주)트루데프 수석연구원

2011년 3월 ~ 현재 삼성전자 무선사업부 책임연구원

※주관심분야: 무선 센서 네트워크, 오디오/비디오 신호처리 및 응용